Bangalore, Karnataka
shreyas.college@gmail.com

Shreyas V

GitHub: shreyasvinaya
LinkedIn: shreyas-v1

Education

BE Computer Science, *Birla Institute of Technology and Science, Pilani – Goa Campus* 2020-2025
MSc Chemistry, *Birla Institute of Technology and Science, Pilani – Goa Campus* 2020-2025

Publications

- **Sathyanarayana, S. V.**, Hiremath, S. D., Shah, R., Panda, R., Jana, R., Singh, R., ... & Ramsundar, B. (2025). "DeepRetro: Retrosynthetic Pathway Discovery using Iterative LLM Reasoning". *Accepted at Nature Scientific Reports*.
- **V. Shreyas**, J. Siguenza, K. Bania & B. Ramsundar "Open Source Molecular Processing Pipeline for Generating Molecules" *Molecular Machine Learning Conference (MoML) 2024, Baylearn 2024, ML4PS workshop at NeurIPS 2024*
- R. Singh, A. A. Barsainyan, R. Irfan, C. J. Amorin, S. He, T. Davis, A. Thiagarajan, S. Sankaran, S. Chithrananda, W. Ahmad, D. Jones, K. McLoughlin, H. Kim, A. Bhutani, **S. V. Sathyanarayana**, V. Viswanathan, J. E. Allen & B. Ramsundar "ChemBERTa-3: An Open Source Training Framework for Chemical Foundation Models." (2025). *Accepted at Royal Society of Chemistry's Digital Discovery*
- Bisoi, A. V., **Shreyas, V.**, Siguenza, J., & Ramsundar, B. "DeepChem-Variant: A Modular Open Source Framework for Genomic Variant Calling". *In Championing Open-source DEvelopment in ML Workshop@ ICML25.*
- Mohanty, D., **Shreyas, V.**, Palai, A., & Ramsundar, B. (2024). Open-source Polymer Generative Pipeline. *4th Annual AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE)*, 2025.
- H. Mestha, K. Bania, **V Shreyas**, S. Liu, A. Srinivasan (2025). Multi-Turn Human–LLM Interaction Through the Lens of a Two-Way Intelligibility Protocol. MTI-LLM Workshop at NeurIPS 2025.
- **V. Shreyas** and A. Swati, "Predicting ATP binding sites in protein sequences using Deep Learning and Natural Language Processing" *3rd Annual AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE)*, 2024.
- H. Mestha, T. Agrawal , K. Bania, **V. Shreyas** & Y. Bhisikar (2024). "CountCLIP–[Re] Teaching CLIP to Count to Ten". *arXiv preprint arXiv:2406.03586.*
- P. Basa Pati and **V. Shreyas**, "Speech to Equation Conversion using a PoE Tagger," *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, 2022, pp. 1-4, doi: 10.1109/I2CT54291.2022.9824252.

Research Experience

Research Fellow(DeepRetro Lead) **Jun 2023 — Aug 2025**
Deep Forest Sciences *Remote*

- Developed tools for **AI-aided retrosynthesis**, **molecule & materials discovery** and built infrastructure to train **Molecular LLMs**
- Led the design and development of **DeepRetro: a Chemist's AI Co-Pilot for Collaborative Retrosynthesis**. DeepRetro is a novel, **open-source hybrid AI framework** that solves complex retrosynthesis problems.
- Architected a core "**iterative LLM reasoning**" loop that integrates the generative power of LLMs with the precision of traditional template-based engines.
- Engineered a **rigorous validation pipeline** to check LLM-generated pathways for chemical validity, stability, and hallucinations, recursively feeding valid precursors back into the search.
- Achieved **state-of-the-art performance**, successfully solving 96.3% of targets on the challenging USPTO-190 multi-step benchmark.
- Demonstrated novel scientific discovery by generating viable, previously **unreported synthetic pathways** for complex natural products like Ohuamine C.
- Migrated models in the open-source drug discovery toolkit **DeepChem** from TensorFlow to **PyTorch**, increasing user base by **92%**

Visiting Researcher Intern **Jun 2024 — Present**
Yale University *New Haven, NY*

- Conducting research on **molecule generation** models using **geometric deep learning** with Prof. Smita Krishnaswamy
- Developing novel algorithms to enhance molecular design efficiency

Research Intern **Dec 2023 — Mar 2025**
Harvard Medical School *Remote*

- Investigating the neural mechanisms of language processing using **AI models** under Prof Gabriel Kreiman
- Applying advanced machine learning techniques to analyze neural data, contributing to a deeper understanding of language cognition

AI Research Scientist **Aug 2025 — Present**
MStack AI *Bangalore, India*

- Designing novel **Chemistry-first AI models** to tackle Molecule Synthesis and Chemical supply chains.
- Led the design and development of **chemical foundational models** for chemical reaction understanding

Student Researcher **2022 — 2024**
APPCAIR, Birla Institute of Technology & Science, Pilani *Goa*

- Researched the use of **Large Language Models** to generate novel molecules for drug discovery and predict their **synthesizability**
- Devised deep learning methods to predict **protein residue binding** sites
- Evaluated **BERT** and **Transformer**-based architectures for protein residue binding location prediction

Skills

Tools and Languages	Python, Git, LaTeX, C, MySQL, PyTorch, TensorFlow, Pandas, CI/CD Pipelines, AWS, Azure
Research Domains	Machine Learning, Drug Discovery, Generative AI, NLP, LLMs, GNNs, Computational Bio-Chemistry

Projects

Metamorph - Generative AI **May 2023 — July 2023**
Building an **emotionally responsive Humanoid** bot with **knowledge capture , speech text generation** using Generative Large Language Models

Activities

Mentor, DeepChem, Google Summer of Code **Summer 2024,2025**

- Mentored Debasish Mohanty and assisted him in adding polymer generative pipelines to the open-source library DeepChem
- Mentored Bibhusundar Mohapatra and helped him port the Chemception model from TensorFlow to PyTorch

ML Lead, Google Developer Student Clubs, BITS Goa **2022 — 2023**

- Guided a group of 25+ ML enthusiasts , conducted weekly ML training sessions
- Organized hackathons , mentored in ML model building , and fostered exploration in ML research.

Awards and Recognitions

Winner ($100,000) *Standard Industries Chemical Innovation Challenge* *Dec 2024*
2nd Runner up | *ACM Goa Chapter's event PitchIt!* *Apr 2024*